1996

Levels of aggregation in survey analysis

Stephenie Anne Posyluzny
San Jose State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

LEVELS OF AGGREGATION IN SURVEY ANALYSIS

A Thesis

Presented to

The Faculty of the Department of Psychology

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

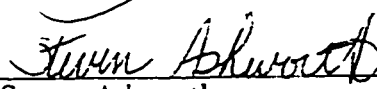
by

Stephenie Anne Posyluzny

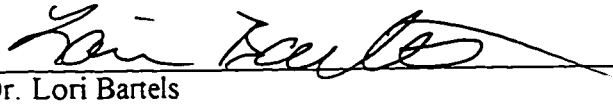
December 1996

APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

Dr. Howard Tokunaga



Dr. Steven Ashworth



Dr. Lori Bartels

APPROVED FOR THE UNIVERSITY

ABSTRACT

LEVELS OF AGGREGATION IN SURVEY RESEARCH

by Stephenie Anne Posyluzny

Abstract

A common approach to studying people within organizations has been to gather whatever data can be collected and perform aggregate analyses. The purpose of the current study is to look at issues surrounding aggregation of employee survey data in order to study the organization.

Studying the effects of variables at one level of an organization on those at a higher organizational level have identified many theoretical issues and has begun to generate new theory. It has been suggested that the problem with aggregation is that researchers are ignoring the unit or level of theory, the level of measurement, and the level of statistical analysis. The current study takes into consideration these three levels and uses intraclass correlation and WABA techniques in order to test the appropriateness of aggregation.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Steve Ashworth, Ph.D. for having the patience to get me through this thesis. I truly learned a lot from Steve and I appreciate him taking time to share some of his knowledge with me. He is a true intellectual and I respect him greatly.

Secondly, I would like to thank Howard Tokunaga, Ph.D. and Lori Bartels, Ph.D. for everything. How do I even begin to tell you all that this education has done for me? You both gave me such a strong knowledge in I/O Psychology and I left SJSU knowing how to apply the things you taught. Lori, you taught me to enjoy the softer side of Organizational Development and Howard, you taught me the more analytical side of Industrial Psychology. The knowledge you both passed on has made me a more marketable person today.

Lastly, to my parents and grandparents. Thank you for all your support and love over the years. Without all of you, I never would have made it this far. I love you with all my heart!

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LEVELS OF AGGREGATION IN SURVEY ANALYSIS

Stephenie A. Posyluzny

San Jose State University

Running head: LEVELS OF AGGREGATION

Abstract

A common approach to studying people within organizations has been to gather whatever data can be collected and perform aggregate analyses. The purpose of the current study is to look at issues surrounding aggregation of employee survey data in order to study the organization.

Studying the effects of variables at one level of an organization on those at a higher organizational level have identified many theoretical issues and has begun to generate new theory. It has been suggested that the problem with aggregation is that researchers are ignoring the unit or level of theory, the level of measurement, and the level of statistical analysis. The current study takes into consideration these three levels and uses intraclass correlation and WABA techniques in order to test the appropriateness of aggregation.

One clear goal of Industrial/Organizational Psychology is the study of people within organizations. This is usually done by studying either the group or the individual. However, many I/O Psychologists are investigating group-level processes to the virtual exclusion of individual-level processes. For example, more and more organizations are restructuring and individuals are becoming members of a team or work group. In studying an organization, one is thus studying a complex system with varying organizational levels. In order for the field of I/O Psychology to grow, it is essential that researchers begin using a multi-level approach to the study of organizations (Rousseau, 1985). The purpose of the current study is to look at issues surrounding aggregation of survey data.

A common approach to studying organizations has been to gather whatever data can be collected and perform aggregate analyses by pooling the data from all individuals in the sample (Steckel, Lehmann & Corfman, 1988). In other words, "aggregation is the combination of information from one level to represent attributes of a higher level unit (from pupils to schools, from departments to organizations)" (Rousseau, 1985, p. 5). Comparing relationships between variables at different levels of analysis (e.g., individual and organizational) requires the collection of data at each of the different levels (Rousseau, 1985). Jennifer George (1990) studied the relationship between personality, affect, and behavior at the work group level of analysis by aggregating individual data. However, oftentimes, researchers do not have global data for the organizational variables of interest. Additionally, because of concerns surrounding confidentiality in surveying, organizations do not want to report individual data. Therefore, researchers tend to rely on aggregated data from individuals to represent the organizational-level variable (Ostroff, 1993).

It has been nearly fifty years since Robinson (1950) warned against the use of aggregate data to study individuals. Since then, several authors have discussed the question of whether and when it is appropriate to aggregate data (Robinson, 1950; Jones & James, 1979; Rousseau, 1985; Ostroff, 1993). Studies which aggregate individual data in order to study global issues have often shown weaker correlations at the individual level and stronger correlations at the organizational level. These correlations at the group or organizational level have been found by Jones and James (1979) in their study of climate perceptions and context, structure, and demographic variables. Schmitt, Colligan, and Fitzgerald (1980) also found strong correlations at the group level in their study of job attitudes and stress. Yet another study by Ostroff (1992) also found strong correlations at the group level between satisfaction and performance. These studies may be giving other researchers the justification to use aggregated data. However, even though these studies found a strong correlation at the organizational level, Ostroff (1993) warns that researchers may be failing to consider random error and measurement error in their interpretations of results. In doing this, researchers could be making "erroneous interpretations about the strength of the relationships among variables at different levels of analysis" (Ostroff, 1993, p. 569).

It has been proposed by Jones and James (1979) that before using aggregated data, some kind of empirical rationalization that aggregation is justifiable is required. It has been suggested that factors such as job type, technology, individual differences, departmental differences, etc. may contribute to heterogeneous perceptions across individuals (Jones & James, 1979). Because of this, aggregation of individual level data may lead to erroneous conclusions. Jones and James (1979, p. 208) suggested the following criteria be used to justify aggregation: "(a) significant differences in aggregated

or mean perceptions across different organizations or subunits; (b) interperceiver reliability or agreement, (c) homogeneous situational characteristics and (d) meaningful relationships between the aggregated score and various organizational, subunit, or individual criteria."

Other researchers have gone a step further and cautioned researchers about several fallacies associated with the use of aggregated data. One fallacy, the fallacy of the wrong level, occurs when correlations at a more macro level are used to make inferences about individuals, or vice versa. Researchers have found that information based on aggregated data provide inflated estimates of relationships at the individual level of analysis (James, 1982). Robinson (1950) showed that an aggregated correlation cannot always be compared. Hence, researchers have been warned not to assume that a relationship among variables at one level represents the same relationship at another level of analysis (Ostroff, 1993).

James (1982) concluded from his research on individual's perceptions of climate that if one is assessing individual data then one needs to analyze the data at the individual level. James provides several examples of research in which individual data was not used in analyses. One such example was Schneider and Bartlett's research on climate perceptions in which the means of agents and agency managers were correlated (James, 1982). James (1982) goes on to state that the high levels of agreement found were most likely inflated because the agent's answers were aggregated with agency manager's answers. By doing this, researchers deleted within-group variance in the agents' perceptions (James, 1982).

In Robinson's study (Firebaugh, 1978) it was shown that correlations between variables at the aggregate level differed from correlations between the same variables at

the individual level. From this finding Robinson concluded that researchers should not use aggregate data to study individuals; those who did were said to be guilty of the "ecological fallacy" (Firebaugh, 1978). In general, the ecological fallacy refers to the role that level itself plays in altering the relationships among variables (Rousseau, 1985).

Yet another fallacy is a "contextual fallacy." A researcher may confront a contextual fallacy when s/he does not specify the effects that social or physical settings (i.e., work group members, technology available, organizational culture, etc.) have on the relationship between variables (Rousseau, 1985). Firebaugh (1978) suggested that analyses that use aggregate data to study individual-level effects can produce a contextual fallacy which can in turn provide the researcher with biased estimates.

Even though researchers have been cautioned, I/O psychologists have continued to use aggregate data to make inferences about individuals. Researchers probably continue to use aggregate data because appropriate individual-level data are often unavailable or because one simply doesn't know of a more appropriate way of looking at the data. Researchers usually find that the pertinent data are only available in aggregated form. Similarly, some researchers may have information on many individuals and wish to say something about the social setting in which these individuals work. Rather than measure something about the settings directly, convenience may dictate the aggregation of the individual data to a higher level to approximate information about the "focal unit" (the unit to which generalizations are made). Either way, there is incongruity among the focal unit level of measurement, and the level of analysis. It is this incongruity and not the process of aggregation that is of concern.

Many researchers have asked, "When is it appropriate to aggregate individual data?" To examine the levels to which the leadership construct may be linked,

Dansereau, Alutto and Yammarino (1984) describe a technique for analyzing leadership effects: within and between analysis (WABA). When individual-level data exist WABA examines the extent of individual and group-level effects by comparing within- and between-eta correlations. However, this technique does not establish the construct validity of the aggregated variable; it merely indicates whether or not there is evidence of a group-level effect on an individual-level variable (Ostroff, 1993).

George (1990) uses the WABA technique and states that in order to test the appropriateness of aggregation, a researcher needs to test the within-setting agreement between subjects. George (1990) used WABA in order to test personality, affect, and behavior as group-level phenomena. In the study, it was hypothesized that characteristic levels of the personality traits *positive affectivity* and *negative affectivity* within groups influence the positive and negative affective tones of the groups (George, 1990). WABA results in George (1990) indicated that the relations between positive affective and negative affective tones of the groups held at the group level of analysis and at the individual level of analysis.

Levels Theory

Recent discussions of multi-level research issues have focused primarily on problems of aggregation (Rousseau, 1985). Studying the effects of variables at one level on those at another have identified many theoretical issues and has begun to generate new theory. However, the study of cross-level effects has been limited to studies on culture and climate, leadership, and work-unit design (Rousseau, 1985).

It may be that the problem surrounding the study of levels and organizations is that researchers need to pay attention to the unit or level of theory, the level of measurement, and the level of statistical analysis (Klein, Dansereau, & Hall, 1994). The

level of theory is important because it "describes the target (e.g., individual, group, organization) that a theorist or researcher aims to depict and explain" (Klein et al., 1994, p 9). It has also been suggested that the level of theory is important because constructs can have completely different meanings at different levels of analysis (Whitley, 1992). "The level of measurement refers to the unit to which the data are directly attached (e.g., self-report data are generally individual level, the number of group members is measured at the group level)" (Rousseau, 1985, p. 4). The level of statistical analysis is "the unit to which the data are assigned for hypothesis testing and statistical analysis" (Rousseau, 1985, p. 4). However, just because one establishes the level of theory doesn't mean there will not be inconsistencies with the level of measurement and/or the level of statistical analysis (Klein et al., 1994).

Roberts, Hulin, and Rousseau (1978) argue that when data are aggregated, the aggregated data is not directly linked to the focal level and thus ambiguity and confusion arise when interpreting the data. The assumption underlying the use of aggregate data to represent higher level or organizational characteristics is that the aggregated variable represents another form of the construct at a higher level of analysis (Rousseau, 1985). When one specifies the group as the level of theory, "one predicts that group members are sufficiently similar with respect to the construct in question that they may be characterized as a whole" (Klein et al., 1994, p. 10).

Klein et al., (1994) describe three levels of theory: homogeneity, independence, and heterogeneity. When researchers specify that the level of theory is homogeneous, they are hypothesizing that the members of a particular group have similar viewpoints or beliefs on the construct of interest (Klein et al., 1994). In proposing that groups are homogenous, "one focuses on variation between groups, positing that differences between

groups on one construct of the theory are related to differences between groups on other constructs of the theory" (Dansereau, Alluto, & Yammarino, 1984). In order to test a theory which specifies homogeneity, it has been suggested that the researcher use a research measure that focuses on the unit as a whole and one which minimizes within-group variability (Klein et al., 1994). In order to measure homogeneity, a researcher might ask questions that would represent individuals as being a part of a group or whole. An example would be: In general, how would the members of your work group rate his/her manager on .....? (Schneider, 1990). To test this level of theory, one would perform analysis with aggregated or group-level scores (Klein et al., 1994).

In specifying that groups are free of group influence or "independent" one is predicting that individual answers are unconstrained by outside influences (Klein et al., 1994). "In conceptualizing the relationship between two constructs at this level, one thus suggests that between-individual variability in one construct is related to between-individual variability in a second construct" (Klein et al., 1994, p. 13). In other words, by specifying independence one would have equal within- and between-group variance. In order to measure independence, a researcher should use measures that present individual differences (Klein et al., 1994). In order to do this a researcher should ask questions which concentrate on an individual's uniqueness. An example of this would be: How would you rate your manager on.....? (Klein, et al., 1994). In order to test the theory as specifying the individual, one would not use aggregated scores, but rather, individual scores as their analysis for testing the level of theory (Klein et al., 1994).

Lastly, when on specifies a group as being heterogeneous, or as individuals within a group, one is focusing on individual attributes relative to the group average for this attribute (Klein et al., 1994). This level of theory, also known as the "frog pond,"

suggests that an "individual's score on Y depends not only on his/her score on X but also on the size of X relative to the other members of the social group" (Firebaugh, 1980, p. 46). In other words, an individual's score may be undifferentiated in one group and yet relatively large in another in another group (Klein et al., 1994). In specifying heterogeneity, one would be creating a norm or zero-sum for the predictor construct of interest and individuals would vary around the group norm (Klein et al., 1994). In order to measure theories which predict heterogeneity, one would want to maximize within-group variability and compare individuals scores on X to the group mean (Klein et al., 1994). In order to do this a question could be asked in the following manner: Compared to the other members of your group, how would you rate your manager on.....? (Schneider, 1990, p. 393). "If the level of theory is the individual within the group, researchers are advised to use deviation scores (or control for between-group differences in some other way) to test the theory" (Klein et al., 1994, p. 33).

There has not been much research testing levels theory. Most research has only been done comparing correlations at the individual level and group levels. Re-wording questions so that they are relevant to the "focal unit" and testing the level of aggregation has not been formally tested, only theorized. It is this multi-level theory which builds the foundation for the current study.

Company of Study

Before explaining the study, it is necessary to explain the organization in the study. A large financial institution, which employs over 45,000 individuals in the United States, with a research office located in the Bay Area, conducts a yearly opinion survey which measures employee attitudes. The employee survey is a tool that was designed to (1) identify employee concerns, (2) encourage upward communications, (3) develop

action plans, (4) improve employee morale, (5) improve organizational performance, and (6) to measure change.

At the end of the survey season, company-wide results are produced for corporate level officers to look at and see what is happening in the company. These results are produced by combining every individual in every unit, in every division, in every area, in every department, in every office, in every business unit. It is important to know if the current reports are accurately representing work groups.

The current study attempts to analyze "what kind" of information is being reported at the organizational level. It attempts to look at the accuracy of the information being provided and whether an office is falling victim to a fallacy of the wrong level. For the purposes of this study, this researcher will be testing levels of theory (homogeneity, independence, and heterogeneity) and analyzing data for individuals and comparing it to aggregated data at the organizational level. The current study looks at new ways to phrase questions by specifying homogeneity or heterogeneity instead of independence. Do questions which specify heterogeneity aggregate better than those which specify independence? What if one work group is significantly lower on satisfaction? Is this office losing the power of its voice because its answers are being aggregated with all the other people in the organization?

Hypotheses

There are three hypotheses in this study which stem from Klein et al. (1994) research on levels theory. They have suggested that when the level of theory has been specified, one can collect data in a manner that ensures the conformity of the data to the appropriate level. Failing to specify the level of theory may cause confusion when drawing conclusions from the data reported. Hypothesis one focuses on the issues of

specifying independence as the level of theory. The second hypothesis focuses on the issues of specifying homogeneity. Hypothesis three focuses on the level of theory being specified as heterogeneous.

Hypothesis 1: The literature regarding questions specifying independence reveals that individual responses are free of group influence (Klein et al., 1994). Thus, the value of a construct for an individual member of a group is independent of the value of the construct for other members of the same group. This level requires you to use measures that draw attention to the individual's uniqueness.

Following from research done by Klein et al. (1994) it is has been suggested that by asking questions which specify independence, one should keep unaggregated, individual scores and not aggregate the data to a group level. It has also been suggested that by aggregating these questions, one is guilty of the fallacy of the wrong level. Because of this, it is hypothesized that questions specifying independence are the least valuable questions to aggregate and that one would be making the between-group variance and the within-group variance equal.

Hypothesis 2: The questions specifying homogeneity are written so that they specify the group as a whole. The literature states that by doing this one would be attempting to predict that group members are similar with respect to the construct in question, so that group members may be characterized as a whole (Klein et al., 1994). Following from this rationale, one would be focusing on the group as a whole and thus it is hypothesized that one would be maximizing between-group variability while minimizing within-group variance.

Because questions that specify homogeneity regard the group as a whole, it is hypothesized that these questions are more appropriate for group level reports and will

aggregate to the organizational level better than questions which specify independence or heterogeneity. By specifying homogeneity one regards group members as a whole and is not guilty of the fallacy of the wrong level when reporting group-level data.

Hypothesis 3: Items which specify heterogeneity were written so that they specify individuals within the group. These items focus on individual attributes relative to the group average for this attribute. It is hypothesized that one would be minimizing between-group variance while maximizing within-group variance.

By specifying heterogeneity, it is predicted that one would be creating a norm quality for the question. When specifying heterogeneity, one is assuming that group members will vary with respect to the theory's construct, however, the group is seen as a meaningful entity and thus one would be minimizing between-group variability. It is further hypothesized that questions specifying heterogeneity would aggregate better than those specifying independence, however, not as good as those specifying homogeneity.

Table 1 presents the question types and the between- and within-group variance predictions according to the previous hypotheses.

Method

Subjects

The data in this research was gathered from a large financial institution, which employs over 45,000 individuals in the United States, with a research office located in the Bay Area. Approximately 7,000 employees took the computerized version of the employee opinion survey. Of those employees, 33% or 2,300 employees were chosen to participate in the experimental section of the survey.

Materials

The survey consisted of the regular employee opinion survey questions plus an

Table 1

Question Types for Between- and Within-Group Variance Predictions

Type of question
Independence
Equal Between and Within-Group
Homogeneity
Maximize Between-Group
Minimize Within-Group
Heterogeneity
Minimize Between-Group
Maximize Within-Group

additional four questions at the end of the survey which pertained to the current study. Two questions, which specified a level of independence, were taken from the core of the employee opinion survey and reworded to test levels of aggregation. The questions specifying independence were reworded so that they specified homogeneity and heterogeneity (See Table 2). Subjects were asked to give a response to each question using a five-point Likert scale ranging from (1) very good to (5) very poor for the first set of questions, and a scale of (1) strongly agree to (5) strongly disagree on the second set of questions.

The frame was added to the questions specifying homogeneity and heterogeneity in order to get employees thinking about their reference group even before they answered the questions.

Procedure

Employees taking the computerized survey were given a user id and a password. The current study used a quasi-random sampling technique (i.e., every employee with a userid divisible by three received the experimental questions).

For the experimental section of the survey, employees were guaranteed their answers would be kept anonymous and confidential. An introduction screen informed them that their answers would not be summarized in any reports and were to be used to improve the quality of the survey. Additionally, employees were informed that their participation was completely voluntary and they could exit the survey at any time.

Results

Levels theorists have urged researchers to align the level of analysis with the level of theory being tested. In other words, if one is analyzing data which specifies independence, one might use unaggregated, individual scores to test the theory. If one is

Table 2

Survey Questions

Specifying Independence

1. Overall, how good a job do you feel is being done by your immediate manager?
2. *Company Name* supports employees in balancing work and personal life responsibilities.

Specifying Homogeneity

1. Answer the following question according to how you think the **people in your work group** would respond. Overall, how good a job do the people in your work group feel is being done by your immediate manager?
2. Answer the following question according to how you think the **people in your work group** would respond. In general, how would the people in your work group rate *company name* in supporting employees in balancing work and personal life responsibilities?

Specifying Heterogeneity

1. For the following question **compare** your feeling to the feelings of those in your work group: How good a job do **you** feel is being done by your immediate manager?
2. For the following question **compare** your feelings to the feelings of those in your work group: How would **you** rate *company name* in supporting employees in balancing work and personal life responsibilities?

---

testing the theory of homogeneity, one may choose global, group-level scores or individual scores aggregated to the group level.

James (1982) states that provided one has randomly selected individuals, the mean squares provided by one-way ANOVA can be used to compute an intraclass correlation. The form of intraclass correlation that was used in the current study was ICC(1,1) (see Table 3 for results). ICC (1,1) provides researchers with an estimate of interrater reliability at the individual level, and is interpreted as the reliability of a single rating or measurement (James, 1982). Results for ICC(1,1) on the first set of questions (questions regarding job being done by immediate manager) revealed the following: (1) question specifying independence, ICC(1,1)=.04, (2) question specifying homogeneity, ICC(1,1)=.05, and (3) question specifying heterogeneity, ICC(1,1)=.05). The second set of questions (questions regarding work and personal life) revealed the following: (1) question specifying independence, ICC(1,1)=.03, (2) question specifying homogeneity, ICC(1,1)=.06, and (3) question specifying heterogeneity, ICC(1,1)=.04. Looking at these results, one might conclude that none of these questions should be analyzed at the individual level, *including* questions specifying independence.

Intraclass correlations can also be used to estimate the reliability of the mean scores. The reliability of means is estimated by the following equation: [ICC(1,1)]/[1 + (n_k - 1)ICC(1,1)] (James, 1982). The reliability of means will be referred to as ICC(1,k). Results for ICC(1,k) on the first set of questions (questions regarding job being done by immediate manager) revealed the following: (1) question specifying independence, ICC(1,k)=.53, (2) question specifying homogeneity, ICC(1,k)=.62, and (3) question specifying heterogeneity, ICC(1,k)=.58. The second set of questions (questions regarding work and personal life) revealed the following: (1) question specifying independence,

ICC(1,k)=.47, (2) question specifying homogeneity, ICC(1,k)=.66, and (3) question specifying heterogeneity, ICC(1,k)=.54.

The results of ICC(1,k) reveal that questions specifying homogeneity aggregate better than those specifying independence or heterogeneity. The results also show that questions specifying heterogeneity aggregate better than those specifying independence. Looking at these results one might also conclude that all of the questions can aggregate, including those specified at the individual level (ICC(1,k)) GE .53. However, one must exercise caution in interpreting the intraclass correlation results. James (1982, p. 222) warns that ICC(1,k) "pertains to the reliability of means, *not* agreement among individuals" and that it is "entirely possible to obtain a trivial ICC(1,1) and a sizable ICC(1,k)." James (1982) states that the reason for this is that "the larger the n_k (number of individuals) the more stable the mean *X* scores. "For example, if n_k =300 and ICC(1)=.05, then ICC(2)=.94. Given a significant *F* ratio, these statistics suggest that the means are highly reliable (stable), that the *K* organization can be reliably differentiated in terms of perceptions on *X*, and that individuals within organizations *do not agree*" (p. 222).

In acknowledging James' warning, further analyses were done to test for the variability of within- and between-group variability and to test for the appropriateness of aggregation. WABA involves the calculation of two types of correlation coefficients- a within-group correlation and a between-group correlation. The between-group correlations are calculated by using group mean scores that are distributed back to individuals in each of the groups; these coefficients represent the relationships across groups (*Model* divided by the square root of *Total* times *Model*). "Within-group correlations control for group differences and emphasize individual-level relationships"

Table 3

Results for Intraclass Correlations

Question type	Satisfaction of job done by immediate manager		Balance between work and personal life	
	ICC (1,1)	ICC (1,k)	ICC (1,1)	ICC (1,k)
Specifying Independence	.04	.53	.03	.47
Specifying Homogeneity	.05	.62	.06	.66
Specifying Heterogeneity	.05	.58	.04	.54

(George, 1990, p. 112). Within-eta is calculated by (*Error* divided by the square root of *Total* times *Error*). WABA allows for the simultaneous demonstration of individual- and group-level effects (George, 1990). Results of WABA analyses can be seen in Table 4.

An individual response was operationally defined as all the individuals who worked in the same work unit. Following from George (1990, p. 110) "rather than measure something about the setting directly, convenience may dictate the aggregation of the individual to a higher level in order to learn pertinent information about the focal unit." As was stated earlier, it is the only incongruity among the focal unit level of measurement and the level of analysis that is of concern, not the actual aggregation. The level of analysis for questions specifying independence, homogeneity, and heterogeneity is in line with the focal level of measurement.

Results show that within-group correlations were all larger than the between-group correlations. The results of the WABA analysis imply that all the questions asked are individual-level relations.

Discussion

The results of this study indicate that research hypothesis one was only partially supported. Intraclass correlation results for reliability of means revealed that questions specifying independence were able to aggregate. However, results were stronger for questions specifying homogeneity and heterogeneity. Additionally, WABA results revealed that between-group and within-group variance were not equal. However, the results did indicate that questions specifying independence had individual-level relations.

The results indicate that research hypothesis two was partially supported. Intraclass correlation results for reliability of means revealed that questions specifying

Table 4

Results for Within and Between Analysis

Question Type	Satisfaction of job done by immediate manager		Balance between work and personal life	
	Between-eta	Within-eta	Between-eta	Within-eta
Specifying Independence	.38	.93	.36	.93
Specifying Homogeneity	.41	.91	.43	.90
Specifying Heterogeneity	.40	.92	.38	.93

homogeneity aggregate better than those specifying independence and heterogeneity. However, WABA results revealed that questions specifying homogeneity did not maximize between-group variance and minimize within-group variance.

Results indicate that research hypothesis three was supported. Intraclass correlation results for reliability of means revealed that questions specifying heterogeneity did aggregate better than those specifying independence, but not better than those specifying homogeneity. WABA results also revealed that questions specifying heterogeneity minimized between group variance and maximized within-group variance.

This was a first step in researching levels theory. Klein et al. (1994) presented the theory that researchers need to align their unit of theory with the level of measurement, and level of statistical analysis. The current study tried to do just that. Unfortunately, due to a relatively small sample size, the data collected for this study are not adequate to directly address all hypotheses. However, results from this study do suggest that there is further research to be done.

One change that could be made is in the questions themselves. Possibly changing the questions to "The people I work with cooperate to get the job done" or "How would you rate the overall quality of work done in your work group?" Both of these questions deal with work group issues even though they are being asked at the individual level. The questions that were asked for the purposes of the current study may have been difficult for people to answer, especially the question about "balance between work and personal life." However, perhaps what the results tell us is that everyone answers questions at the individual level. Maybe the current study created its own new fallacy, which is the "fallacy to change an individual's answer." Even though employees were presented with a different frame or way of thinking, they may still be rationalizing and answering

questions according to their won beliefs. Perhaps it is not truly possible to change any question and get an individual to answer each one differently.

Further studies may want to avoid a random selection of individuals within an organization and instead randomly select intact work groups. Because the current study randomly selected employees within an organization this affected the sample size of intact work groups, thus possibly affecting analyses and conclusions. By selecting individuals from intact groups, it would be easier to look at individual responses within a work group. It would also make it easier to then look at differences in the way employees in a work group answer the same questions that specify independence, homogeneity or heterogeneity. With sampling entire work groups, one may obtain more valuable and significant information.

Another issues in the current study is that employees were sampled from organizational offices, further causing relatively small sample sizes in some work groups. Because of this, data was then analyzed at the organizational level and not at the work group level. By sampling according work group, the number of responses by work group could be increased and the analyses could then be done at the work group level. If analysis were done at the work group level a researcher could then better compare responses to test the appropriateness of aggregation to the level of the work group and organizational levels. In doing this, one may find more significant results than the current study.

In order to avoid a source of error, the current study did not change the response categories for each type of question. For instance, the question "how good a job is being done by the immediate manager" has a response scale of "very good" to "very poor." This response scale was used for each type of question (independence, homogeneity, and

heterogeneity). Future studies may want to test the differences in questions when the scales are changed. So that a question specifying heterogeneity would have a scale that is "much better" to "much worse." In questions that specify heterogeneity the questions ask individuals to compare their feelings to those in their work groups. It makes sense to give them a more appropriate scale and then test the appropriateness of aggregation.

In closing, although the results of this study were unable to provide empirical support for all research hypotheses, they suggest that there may be reason look further into the study of levels theory. When researchers study employees, they target a "group" who they are aiming to explain. It should follow then that researchers should align that unit level with a level of measurement and statistical analysis. All the research found to support the current study researched overall attitudes or ideas like satisfaction and performance (Ostroff, 1992) or climate perceptions and context, structure, and demographic variables (Jones & James, 1979) or personality, affect, and behavior (George, 1990), instead of studying individual questions. There is obviously a lot more that researchers can study to find the correct answer of "to aggregate or not to aggregate."

References

Dansereau, F. Alutto, J.A., & Yammarino, F.J. (1984). Theory testing in organizational behavior: The variant approach. Englewood Cliffs, NJ: Prentice-Hall.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. American Sociological Review, 43, 557-572.

Firebaugh, G. (1980). Groups as contexts and frogponds. In K.H. Roberts & L. Burnstein (Eds.), Issues in aggregation (pp. 43-52). San Francisco: Jossey-Bass.

George, J.M. (1990). Personality, affect, and behavior in groups. Journal of Applied Psychology, 75, 107-116.

James, L.R. (1982). Aggregation bias in estimates of perceptual agreement. Journal of Applied Psychology, 67, 219-229.

Jones, A.P. & James, L.R. (1979). Psychological climate: Dimensions and relationships of individual and aggregated work environment perceptions. Organizational Behavior and Human Performance, 23, 201-250.

Klein, K.J., Dansereau, F., & Hall, R.J. (in press). Levels issues in theory development, data collection, and analysis. Academy of Management Review.

Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. Journal of Applied Psychology, 77 (6), 963-974.

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. Journal of Applied Psychology, 78(4), 569-582.

Roberts, K.H., Hulin, C.L., & Rousseau, D.M. (1978). Developing an interdisciplinary science of organizations. San Francisco: Jossey-Bass.

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. American Sociological Review, 15, 351-357.

Rousseau, D.M. (1985). Issues of level in organizational research: Multi-level and cross- level perspectives. Research in Organizational Behavior, 7, 1-37.

Schmitt, N., Colligan, M.J., & Fitzgerald, M. (1980). Unexplained physical symptoms in eight organizations: Individual and organizational analyses. Journal of Occupational Psychology. 53, 305-317.

Schneider, B. (1990). The climate for service: An application of the climate construct. In B. Schneider (Ed.), Organizational climate and culture, 393-412. San Francisco: Jossey-Bass.

Steckel, J.H., Lehmann, D.R., & Corfman, K.P. (1988). Estimating probabilistic choice models from sparse data: A method and an application to groups. Psychological Bulletin, 103, 131-139.

Whitley, B.E. (1992). Units of analysis, measurement scales, and statistics: A comment on Kerwin and Shaffer. Personality and Social Psychology Bulletin, 18(6), 684.